The Bell System Technical Journal
Vol. 51, No. 6, July-August, 1972
Printed in U.S.A.

Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords

By K. ISHIZAKA and J. L. FLANAGAN

(Manuscript received January 13, 1972)

A model of voiced-sound generation is derived in which the detailed acoustic behavior of the human vocal cords and the vocal tract is computed. The vocal cords are approximated by a self-oscillating source composed of two stiffness-coupled masses. The vocal tract is represented as a bilateral transmission line. One-dimensional Bernoulli flow through the vocal cords and plane-wave propagation in the tract are used to establish acoustic factors dominant in the generation of voiced speech. A difference-equation description of the continuous system is derived, and the cord-tract system is programmed for interactive study on a DDP-516 computer. Sampled waveforms are calculated for: acoustic volume velocity through the cord opening (glottis); glottal area; and mouth-output sound pressure. Functional relations between fundamental voice frequency, subglottal (lung) pressure, cord tension, glottal area, and duty ratio of cord vibration are also determined.

Results show that the two-mass model duplicates principal features of cord behavior in the human. The variation of fundamental frequency with subglottal pressure is found to be 2 to 3 Hz/cm H_2O, and is essentially independent of vowel configuration in the programmed tract. Acoustic interaction between tract eigenfrequencies and glottal volume flow is strong. Phase difference in motion of the cord edges is in the range of 0 to 60 degrees, and control of cord tension leads to behavior analogous to chest/falsetto conditions in the human. Phonation-neutral, or rest area of cord opening, is shown to be a critical factor in establishing self-oscillation. Finally, the complete synthesis system suggests an efficient, physiological description of the speech signal, namely, in terms of subglottal pressure, cord tension, rest area of cord opening, and vocal-tract shape.

I. GENERATION OF VOICED SOUNDS IN SPEECH

The vocal cords constitute the sound source for all voiced sounds in speech. The cords consist of opposing ligaments which form a con-

striction at the top of the trachea where it joins to the lower vocal tract. When air is expelled at sufficient velocity through this orifice (the glottis), the cords vibrate and act as an oscillating valve which interrupts the air flow into a series of pulses. These pulses of volume flow serve as the excitation source for the vocal tract in all voiced sounds, and the passive resonances of the vocal tract are excited by the glottal pulses. Voice quality and prosodic features of speech are therefore strongly dependent upon the properties of cord vibration.

In the synthesis of speech by machines (for automatic voice response from computers, for example) it is desirable to make the synthetic voice as natural sounding as possible. Toward this end, it is necessary to understand the fundamental acoustic principles of voiced-sound generation and how these factors might be incorporated into a machine voice. Further, in a rather different area, the successful medical diagnosis (and correction) of voice disorders depends upon an understanding of the critical parameters of vocal-cord behavior. As in the case of computer synthesis, medical diagnosis can be facilitated through an accurate and viable model of the human vocal cords. Applications such as these, together with fundamental interests in the acoustics of speech, provide a motivation for modeling the acoustic behavior of the vocal cords.

II. SELF-OSCILLATING MODELS OF THE VOCAL CORDS

The first quantitative self-oscillating model of the vocal cords was devised by one of the authors and implemented with a vocal-tract synthesizer on a digital computer.[1,2] This model was subsequently elaborated to include the mechanism of voiceless sound generation as well,[3] and was used for the synthesis of simple speech samples.

In this early work, the vocal cords were approximated as a simple mechanical oscillator, composed of single opposing masses, springs, and nonlinear damping–that is, a so-called one-mass approximation of each vibrating cord. The oscillating masses were permitted only lateral displacement and were driven by a function of the subglottal pressure and the Bernoulli pressure in the glottal orifice. The heretofore much-used assumption of linear separability of sound source and vocal tract was not made, and acoustic factors such as voice pitch, waveform of glottal flow, and glottal duty factor were derived as self-determined functions of physiological parameters, namely, subglottal (lung) pressure, vocal-cord tension (or natural frequency), "neutral" glottal area, and vocal-tract shape.

The waveforms of glottal area and volume velocity obtained in this first study were similar to those observed in high-speed motion pictures of the human vocal cords and in inverse filtering of natural speech. Further, the results revealed how the acoustic interaction between the vocal cords and the vocal-tract shape (through its driving-point impedance) could influence the waveform and period of the glottal flow. Control of the physiological parameters, subglottal pressure, cord tension, neutral area, and vocal-tract shape, were shown to be sufficient for the synthesis of voiced and voiceless sounds.[3]

Although the one-mass model could produce acceptable voiced-sound synthesis and simulate many of the properties of glottal flow, it was inadequate to produce other physiological detail in vocal cord behavior. For example, the amount of acoustic interaction displayed between source and tract was greater than observed in human speech.* The one-mass model was congenitally incapable of sustained oscillation for a capacitive input load of the vocal tract–corresponding to oscillation at a frequency just above a formant (or eigen) frequency of the tract. Also, a physiologically-natural correlate of chest and falsetto registers and a phase-difference in the motion of the cord edges were lacking.

To incorporate more physiological properties, multiple-mass representations of the cords were therefore considered.[4–6] The cord ligaments can be mechanically represented with as distributed a system as desired, i.e., periodic structures of masses, springs, and losses. However, theoretical work has shown that a two-mass approximation[6,7] can account for most of the relevant glottal detail, including phase differences of upper and lower edges and oscillation for a capacitive input impedance of the vocal tract. An initial effort at computer simulation of these factors[4] produced realistic phase differences and chest-falsetto dichotomy, but nonrealistic dependence on acoustic load. The difficulty lay in the equivalent circuit of the glottal orifice, the manner of its control, and the physiological data available for the simulation.

The present work seeks a comprehensive and definitive treatment of the relevant acoustic theory and the existing physiological data. As in the earlier study,[2] simulation on an interactive DDP-516 laboratory computer is the tool by which the model is assessed and the unknown constants are estimated. In the sequel, the level of understanding and the realism attained by the two-mass model will be discussed.

---

* The amount of interaction is critically dependent upon the trans-glottal pressure distribution. In the first work, van den Berg's measurements of glottal pressure were used.

III. MECHANICAL RELATIONS FOR THE TWO-MASS MODEL

The vocal cords are assumed to be bilaterally symmetric. The properties of only one cord are therefore discussed, the same being implied for the opposing cord. A schematic diagram of the glottal system is shown in Fig. 1. The trachea, leading to the lungs, is represented by the pipe to the left. The larynx tube, leading to the vocal tract, is to the right. These tubes are assumed to be cylindrical in shape and are fixed in size. The glottis constitutes a constriction between these tubes, and the size of the constriction depends upon the cord displacement. The inlet to the glottal constriction occurs over the contraction distance l_c. Expansion back to the vocal-tract cross section occurs over the distance l_e. Aerodynamic pressures relevant to the following discussion are indicated in Fig. 1.

In the two-mass model, the vocal cord is divided in depth (thickness) into an upper and a lower part. Each part consists of a simple mechanical oscillator having a mass, spring, and damping (m, s, and r). The two masses of a cord, m_1 and m_2 are permitted only lateral motion, x_1 and x_2, and the masses are coupled by a linear spring, of stiffness k_c, as shown in Fig. 1. Other factors shown in Fig. 1 are:

l_g	the effective length of the vocal cords (or of the glottal slit),
d_1 and d_2	the thickness of m_1 and m_2, respectively,
s_1 and s_2	the equivalent springs,
r_1 and r_2	the equivalent viscous resistances,
A_{g01} and A_{g02}	the cross-sectional areas of the glottal, slit when m_1 and m_2 are at rest (i.e., the "phonation neutral" areas),
U_g	the average volume velocity across the glottal area.

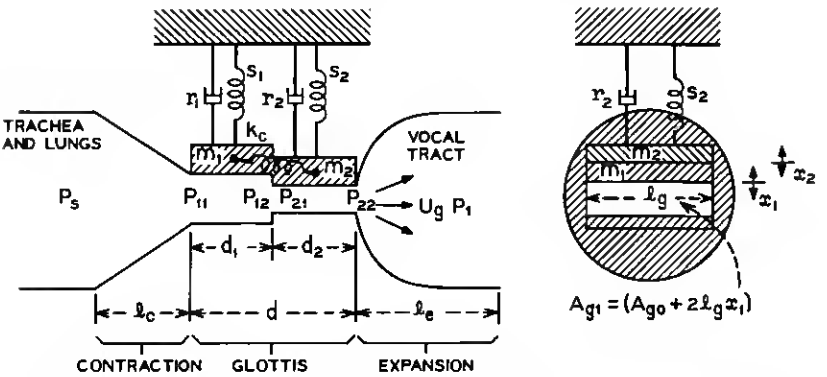

Fig. 1—Schematic diagram of the two-mass approximation of the vocal cords.

Owing to the assumption of bilateral symmetry, variations in cross-sectional areas due to the lateral displacements x_1 and x_2 are doubled to give the total area variation; that is, the cross-sectional areas at the two masses are:

$$A_{g1} = A_{g01} + 2l_g x_1$$
$$A_{g2} = A_{g02} + 2l_g x_2.$$

3.1 *Nature of the Vocal-Cord Springs*

The function of the linear coupling spring, k_c, is to represent an effect of flexural stiffness in the lateral direction of the vocal cords. This variable flexural stiffness results from varying the thickness and stiffness of the cords by action of the thyroarytenoid muscle (vocalis).

The springs, s_1 and s_2, are an equivalent representation of the tension of the vocal cords, which becomes firmer due to contraction of the anteriol cricothyroid muscle and other muscles. The springs, s_1 and s_2, are given a nonlinear characteristic to conform to the stiffness as measured on fresh, excised human vocal cords.[8] The nonlinear relation between the deflection from the equilibrium position and the force required to produce this deflection is given by

$$f_{sj} = k_j x_j (1 + \eta_{kj} x_j^2), \qquad j = 1, 2, \tag{1}$$

where f_{sj} is the force required to produce x_j, k_j is the linear stiffness, and η_{kj} is the coefficient describing the nonlinearity of the spring, s_j, being positive in this case.

During closure of the glottis, the model should satisfy realistic conditions at the colliding surfaces of the vibrating masses, m_1 and m_2 with their opposing counterparts. A contact force at collision will cause some deformation in the flesh of the vocal cords. The restoring force at this deformation can be represented by an equivalent spring s_{hj} ($j = 1, 2$). For simplicity, a nonlinear characteristic, similar to eq. (1), is assumed for the spring s_{hj}, that is,

$$f_{hj} = h_j\left(x_j + \frac{A_{g0j}}{2l_g}\right)\left\{1 + \eta_{hj}\left(x_j + \frac{A_{g0j}}{2l_g}\right)^2\right\} \tag{2}$$

for

$$x_j + A_{g0j}/2l_g \leqq 0 \qquad j = 1, 2,$$

where f_{hj} is the force required to produce the deformation to mass, m_j, during collision, h_j is the linear stiffness, and η_{hj} is a positive coefficient representing the nonlinearity of the contacting vocal cords. The resultant restoring force acting on m_j during closure is, therefore, the sum of f_{sj} and f_{hj}, that is, eq. (1) and eq. (2). This change in spring stiffness at closure is schematically illustrated in Fig. 2.

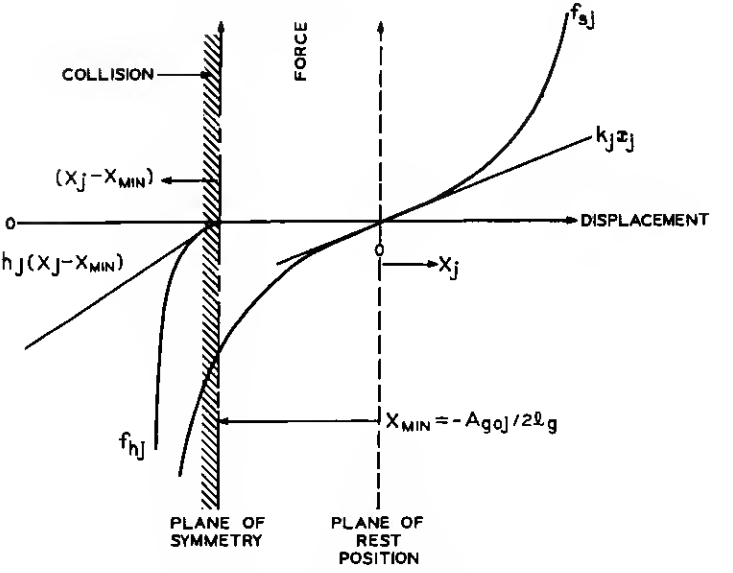

Fig. 2—Characteristics of the nonlinear stiffnesses.

3.2 *Nature of the Vocal-Cord Losses*

As in the earlier formulation,[1] the viscous loss of the vibrating cords is assumed piece-wise linear. The loss is caused to increase step-wise on closure of the cords to represent the "stickiness" of the soft, moist contacting surfaces as they form together.

It is convenient to express the equivalent viscous resistances in terms of damping ratios, ζ_1 and ζ_2, for the uncoupled oscillators, where

$$r_1 = 2\zeta_1\sqrt{m_1 k_1} \quad \text{and} \quad r_2 = 2\zeta_2\sqrt{m_2 k_2}, \tag{3}$$

and where, as shown in eq. (1), k_1 and k_2 are the linear components of stiffness for the springs s_1 and s_2. During the open-glottis condition, the loss is taken as $\zeta_1 = 0.1$ and $\zeta_2 = 0.6$ for a typical condition of the cord model. As in the earlier work, the loss during the closed-glottis condition is taken essentially as critical damping, namely

$$\zeta_1 = (1.0 + 0.1) \quad \text{and} \quad \zeta_2 = (1.0 + 0.6). \tag{4}$$

IV. PRESSURE DISTRIBUTION ALONG THE GLOTTIS

Because of the small dimensions of the glottis (compared to a wavelength at the frequencies of interest), and because of the high velocity

of the glottal flow (compared to the vocal-cord velocity), we can assume the glottal flow to be quasi-steady.[7] We shall use the Bernoulli equation for one-dimensional flow to obtain the pressure distribution along the glottal flow.

The abrupt contraction in cross-sectional area at the inlet to the glottis produces a *vena contracta* surrounded by stagnant air. The *vena contracta* makes the inlet area A_{g1} appear smaller than it actually is and the pressure drop greater than that dictated by an ideal area change. The loss factor for such a contraction has been studied in fluid flow experiments[9] and found to be on the order of 0.4 to 0.5. Flow measurements by van den Berg, et al.,[10] on plaster cast models of the larynx set the loss figure at 0.37. This latter figure is therefore taken to estimate the pressure drop at the inlet, and we fix this drop at

$$P_{B1}(1.00 + 0.37), \quad \text{or} \quad 0.69\rho(U_g^2/A_{g1}^2),$$

where $P_{B1} = \frac{1}{2}\rho u_{g1}^2$ is the Bernoulli pressure, ρ the air density, and u_{g1} the particle velocity at the lower cord-edge.

Within the constriction formed by the lower part of the cord, the pressure drop is assumed to be governed by viscous loss, which is also consistent with van den Berg's measurements. In this region the pressure falls linearly with distance according to a resistance to the volume flow equal to $12\mu d_1 l_g^2/A_{g1}^3$, where μ is the shear viscosity coefficient.

At the junction between the masses m_1 and m_2, the volume flow U_g is continuous, but the particle velocity changes. There is a corresponding abrupt change in pressure equal to the change in kinetic energy per unit volume of the fluid. This pressure change at the junction is

$$\begin{aligned} \Delta p &= 1/2\rho(u_{g1}^2 - u_{g2}^2) \\ &= 1/2\rho U_g^2(1/A_{g2}^2 - 1/A_{g1}^2). \end{aligned} \tag{5}$$

Throughout the constriction formed by the upper cord-edge, m_2, viscous loss is assumed to govern the pressure drop and, like the lower cord portion, the resistance is taken as $(12\mu d_2 l_g^2/A_{g2}^3)$.

At the abrupt expansion of the glottal outlet, the pressure recovers toward the atmospheric value (assuming no constrictions in the relatively large vocal tract). Van den Berg, in his model flow measurements, found the pressure recovery to be about 0.5 P_B. However, for small constrictions this measurement is difficult and uncertain. It seems preferable to base an estimate of the pressure recovery on momentum considerations, which hold in the theory of fluid flow.

Consider at the sudden expansion the relations for Newton's law,

$f = (d/dt)(mv)$. Then, because U_g is continuous,

$$\rho U_g(u_1 - u_{g2}) = A_1(P_{22} - P_1)$$

or

$$\begin{aligned}(P_1 - P_{22}) &= 1/2\rho u_{g2}^2[2N(1 - N)] \\ &= 1/2\rho \frac{U_g^2}{A_{g2}^2}[2N(1 - N)] \\ &= P_{B2}[2N(1 - N)], \end{aligned} \tag{6}$$

where $N = A_{g2}/A_1$, $P_{B2} = 1/2\rho u_{g2}^2$, and A_1 is the input area to the vocal tract. The value of $2N(1 - N)$ is typically in the order of 0.05 to 0.40, which is somewhat smaller than van den Berg's value. This difference is significant to the acoustic interaction between the vocal tract and the cord source.[1] The pressure distribution along the steady flow through the glottis is indicated in Fig. 3.

In the time-varying condition of the cords, the inertance of the air masses involved should also be taken into account. When combined with the loss terms just discussed, the pressure distribution along the glottis is described by

$$\begin{aligned}
P_s - P_{11} &= 1.37 \frac{\rho}{2}\left(\frac{U_g}{A_{g1}}\right)^2 + \int_0^{l_c} \frac{\rho}{A_c(x)}\,dx \cdot \frac{dU_g}{dt} \,* \\
P_{11} - P_{12} &= 12 \frac{\mu l_g^2 d_1}{A_{g1}^3} \cdot U_g + \frac{\rho d_1}{A_{g1}} \cdot \frac{dU_g}{dt} \\
P_{12} - P_{21} &= \frac{\rho}{2} U_g^2\left(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2}\right) \\
P_{21} - P_{22} &= 12 \frac{\mu l_g^2 d_2}{A_{g2}^3} U_g + \frac{\rho d_2}{A_{g2}} \cdot \frac{dU_g}{dt} \\
P_{22} - P_1 &= -\frac{\rho}{2}\left(\frac{U_g}{A_{g2}}\right)^2 \cdot 2 \frac{A_{g2}}{A_1}\left(1 - \frac{A_{g2}}{A_1}\right).
\end{aligned} \tag{7}$$

V. EQUIVALENT CIRCUIT FOR THE GLOTTIS

On the basis of the pressure difference relations of eq. (7), the acoustic impedance elements of the glottal orifice constitute the equivalent circuit shown in Fig. 4, where the U_g current is continuous. The elements

* The $(U_g(dL/dt))$ term in $(d/dt)(LU_g)$ is negligible, where $L = (\rho d/A)$.

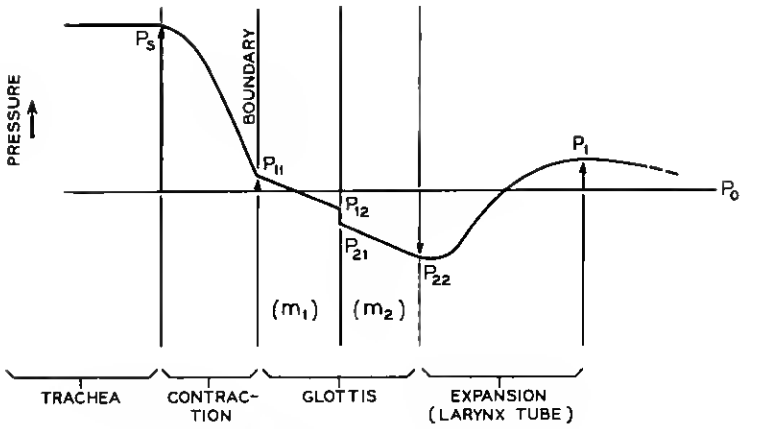

Fig. 3—Pressure distribution along the glottal flow.

of the acoustic circuit are given by

$$R_c = 1.37 \frac{\rho}{2} \frac{|U_g|}{A_{g1}^2}, \qquad L_c = \int_0^{l_c} \frac{dx}{A_c(x)}$$

$$R_{v1} = 12 \frac{\mu l_g^2 d_1}{A_{g1}^3}, \qquad L_{g1} = \frac{\rho d_1}{A_{g1}}$$

$$R_{12} = \frac{\rho}{2} \left(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2} \right) |U_g|$$

$$R_{v2} = 12 \frac{\mu l_g^2 d_2}{A_{g2}^3}, \qquad L_{g2} = \frac{\rho d_2}{A_{g2}}$$

$$R_e = -\frac{\rho}{2} \cdot \frac{2}{A_{g2} A_1} \left(1 - \frac{A_{g2}}{A_1} \right) |U_g|. \tag{8}$$

The total acoustic impedance of the glottis, Z_g, is therefore

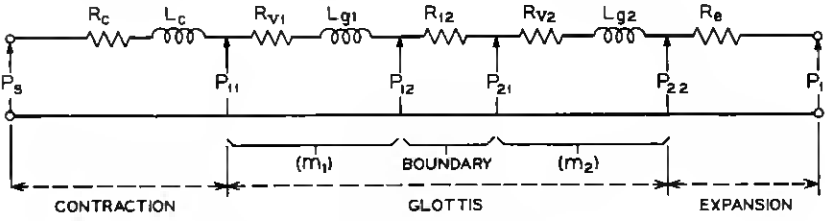

Fig. 4—Equivalent circuit for the glottis.

$$Z_g = \frac{\rho}{2}\,|U_g|\left\{\frac{0.37}{A_{g1}^2} + \frac{1 - 2\,\frac{A_{g2}}{A_1}\left(1 - \frac{A_{g2}}{A_1}\right)}{A_{g2}^2}\right\} + (R_{v1} + R_{v2}) + j\omega(L_{g1} + L_{g2} + L_c) \qquad (9)$$

or

$$Z_g = (R_{k1} + R_{k2})\,|\,U_g\,| + (R_{v1} + R_{v2}) + j\omega(L_{g1} + K_{g2} + L_c),$$

$$Z_g = (R_{k1} + R_{k2})\,|\,U_g\,| + (R_{v1} + R_{v2}) + j\omega(L_{g1} + L_{g2} + L_c), \qquad (10)$$

where

$$R_{k1} = \frac{0.19\rho}{A_{g1}^2}, \qquad R_{k2} = \frac{\rho\left[0.5 - \frac{A_{g2}}{A_1}\left(1 - \frac{A_{g2}}{A_1}\right)\right]}{A_{g2}^2}.$$

In general, L_c can be neglected in comparison to $(L_{g1} + L_{g2})$.

The glottal impedance relation of eq. (10) can be linked to that obtained in flow measurements by van den Berg et al.[10] Using the pressure recovery found by van den Berg for the glottal outlet, namely $1/2\ P_{B2}$, [instead of the momentum relations in eq. (6)] gives a value $R_e = -(\rho/4)\ |\,U_g\,|/A_{g2}^2$. For the case of $A_{g1} = A_{g2} = A_g$, the total glottal impedance becomes

$$z_g = -0.87\,\frac{\rho}{2}\,\frac{|U_g|}{A_g^2} + 12\,\frac{\mu l_g^2 d}{A_g^3} + j\omega L_g\,. \qquad (11)$$

The real part of this impedance is identical with that given by van den Berg.

VI. MODEL SYSTEM FOR VOICED SOUNDS

A network representation of the vocal system for voiced sounds is shown in Fig. 5. Beginning at the left, the subglottal system–comprised

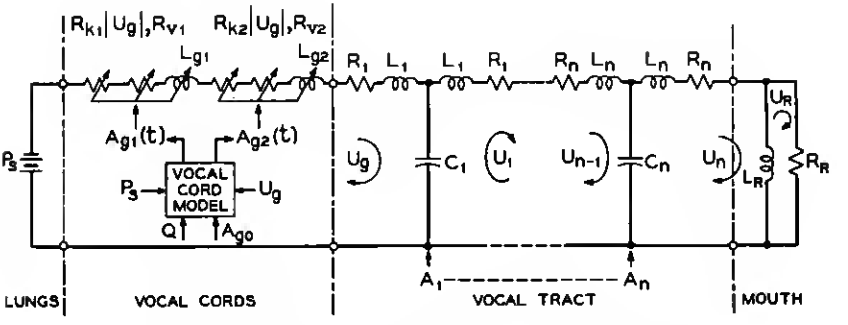

Fig. 5—Network model for the synthesis of voiced sounds.

of the trachea, bronchi, and lungs–is neglected and, as in the earlier study,[1] the subglottal pressure is approximated by a constant excess pressure in the lungs. Neglecting the subglottal system is also based on the finding that its first resonance is relatively high, with a mean value of 650 Hz and a bandwidth of 250 Hz. These values were determined from direct measurements of the subglottal driving-point impedance made on five laryngectomized subjects.[11] The 650 Hz figure is considerably higher than the value of 300 Hz reported by van den Berg.

The vocal tract is represented in Fig. 5 as a transmission line of n cylindrical, hard-walled sections, the element values of which are determined by the cross-sectional areas $A_1 \cdots A_n$, and the cylinder lengths $l_1 \cdots l_n$.[12] The inductances are $L_n = \rho l_n/2A_n$ and the capacitances are $C_n = (l_n \cdot A_n/\rho c^2)$, where c is the sound velocity. In the present work $n = 4$.

To account in part for tract losses, serial resistances R_n are taken to have the form of a viscous loss at the pipe wall, namely $R_n = (S_n/A_n^2)\sqrt{\rho\mu\omega/2}$, where S_n is the circumference of the nth section and ω is the radian frequency. The frequency for evaluation of this loss is fixed at the natural frequency of the lower oscillator, $f = (1/2\pi)\sqrt{k_1/m_1}$, and a multiplicative coefficient (ATT) is applied to increase the loss beyond that contributed by viscous loss at the walls and to produce formant bandwidths appropriate to a closed-glottis condition.* (The typical range for ATT is 20 to 25.)

The transmission line is terminated in a radiation load equal to that for a circular piston in an infinite baffle, namely $L_R = 8\rho/3\pi\sqrt{\pi A_n}$ and $R_R = (128\rho c/9\pi^2 A_n)$, where A_n is the final (mouth) area.[12]

From Fig. 5, the differential equations which relate the volume velocities of the system are:

$$(g\text{-loop}) \quad (R_{k1} + R_{k2})\,|U_g|\,U_g + (R_{v1} + R_{v2})U_g + (L_{g1} + L_{g2})\frac{dU_g}{dt} + L_1\frac{dU_g}{dt} + R_1 U_g + \frac{1}{C_1}\int_0^t (U_g - U_1)\,dt - P_s = 0$$

$$(1\text{-loop}) \quad (L_1 + L_2)\frac{dU_1}{dt} + (R_1 + R_2)U_1 + \frac{1}{C_2}\int_0^t (U_1 - U_2)\,dt + \frac{1}{C_1}\int_0^t (U_1 - U_g)\,dt = 0$$

* Other vocal tract losses not included *per se* are those arising from non-rigid walls and from heat conduction losses at the wall. The former is quite significant in lower-formant damping. The latter is essentially negligible. See Ref. 12.

$$\text{(2-loop)}\quad (L_2 + L_3)\frac{dU_2}{dt} + (R_2 + R_3)U_2 + \frac{1}{C_3}\int_0^t (U_2 - U_3)\,dt + \frac{1}{C_2}\int_0^t (U_2 - U_1)\,dt = 0$$

$$\text{(3-loop)}\quad (L_3 + L_4)\frac{dU_3}{dt} + (R_3 + R_4)U_3 + \frac{1}{C_4}\int_0^t (U_3 - U_L)\,dt + \frac{1}{C_3}\int_0^t (U_3 - U_2)\,dt = 0$$

$$\text{(4-loop)}\quad (L_4 + L_R)\frac{d(U_L)}{dt} + R_4 U_L - L_R\frac{d(U_R)}{dt} + \frac{1}{C_4}\int_0^t (U_L - U_3)\,dt = 0$$

$$\text{(5-loop)}\quad L_R\frac{d}{dt}(U_R - U_L) + R_R \cdot U_R = 0. \tag{12}$$

VII. FORCING RELATIONS FOR THE VOCAL-CORD OSCILLATOR

The masses of the cord oscillator are driven by mean pressures acting on their exposed faces, namely,

$$P_{m1} = \tfrac{1}{2}(P_{11} + P_{12}) = P_s - 1.37\frac{\rho}{2}\left(\frac{U_g}{A_{g1}}\right)^2 - \frac{1}{2}\left(R_{v1}U_g + L_{g1}\frac{dU_g}{dt}\right)$$

and

$$P_{m2} = \tfrac{1}{2}(P_{21} + P_{22}) = P_{m1} - \tfrac{1}{2}\left\{(R_{v1} + R_{v2})U_g + (L_{g1} + L_{g2})\frac{dU_g}{dt}\right\} - \frac{\rho}{2}U_g^2\left(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2}\right). \tag{13}$$

The exposed areas are $l_g d_1$ and $l_g d_2$, respectively. A shape of the cords is assumed such that the forces F_1 and F_2 acting on m_1 and m_2 over their displacements x_1 and x_2 are:

x_1	x_2	$F_1/l_g\,d_1$	$F_2/l_g\,d_2$
$x_1 > x_{1\min}$	$x_2 > x_{2\min}$	P_{m1}	P_{m2}
$x_1 \leqq x_{1\min}$	$x_2 > x_{2\min}$	P_s	0
$x_1 > x_{1\min}$	$x_2 \leqq x_{2\min}$	P_s	P_s
$x_1 \leqq x_{1\min}$	$x_2 \leqq x_{2\min}$	P_s	0,

(14)

where $x_{1\min} = -(A_{g01}/2l_g)$, $x_{2\min} = -(A_{g02}/2l_g)$, and A_{g01}, A_{g02} are the "phonation neutral" values of the glottal area. The equations of motion for the two masses are therefore:

$$m_1 \frac{d^2x_1}{dt^2} + r_1 \frac{dx_1}{dt} + s_1(x_1) + k_c(x_1 - x_2) = F_1$$

$$m_2 \frac{d^2x_2}{dt^2} + r_2 \frac{dx_2}{dt} + s_2(x_2) + k_c(x_2 - x_1) = F_2 ,$$

where

$$A_{g1} = (A_{g01} + 2l_g x_1), \qquad A_{g2} = (A_{g02} + 2l_g x_2),$$

$$s_j(x_j) = k_j(x_j + \eta_{kj} \cdot x_j^3), \quad j = 1, 2, \quad \text{for} \quad x_j > -\frac{A_{g0j}}{2l_g} ,$$

and

$$s_j(x_j) = k_j(x_j + \eta_{kj} \cdot x_j^3) + h_j\left\{\left(x_j + \frac{A_{g0j}}{2l_g}\right) + \eta_{hj}\left(x_j + \frac{A_{g0j}}{2l_g}\right)^3\right\} ,$$

$$\text{for} \quad x_j \leqq -\frac{A_{g0j}}{2l_g} , \qquad (15)$$

and F_1 and F_2 are given by the force table of eq. (14). These equations are coupled to the flow equations through the fact that x_1 and x_2 determine A_{g1} and A_{g2}. Also note that the coupling between the masses, which are permitted only lateral motion, has been linearized to be proportional to $(x_2 - x_1)$. [A more detailed consideration of the elongation produced in the coupling spring by a displacement difference $(x_2 - x_1)$, and of the lateral component of restoring force, leads to modifying the coupling term to $2k_c(x_2 - x_1)^3/(d_1 + d_2)^2$.]

VIII. DIGITAL SIMULATION

The differential equations are approximated by difference equations in which

$$\frac{df(t)}{dt} \cong \frac{f(t_i) - f(t_{i-1})}{(t_i - t_{i-1})} = \frac{f_i - f_{i-1}}{T}$$

$$\int_0^t f(t)\, dt \cong (t_i - t_{i-1}) \sum_{j=0}^{i-1} f(t_j) = T \sum_{i=0}^{i-1} f_i . \qquad (16)$$

These discrete approximations applied to eqs. (12) and (15) yield:

$$(g\text{–loop}) \quad (R_{k1i} + R_{k2i})\,|U_{gi}|\,U_{gi} + (R_{v1i} + R_{v2i})U_{gi}$$
$$+ (L_{g1i} + L_{g2i})\,\frac{U_{gi} - U_{gi-1}}{T} + L_1\,\frac{U_{gi} - U_{gi-1}}{T}$$
$$+ R_1 \cdot U_{gi} + \frac{T}{C_1}\sum_{j=0}^{i-1}(U_{gj} - U_{1j}) - P_s = 0$$

$$(1\text{–loop}) \quad \left\{\frac{L_1 + L_2}{T} + (R_1 + R_2)\right\} U_{1i} - \frac{L_1 + L_2}{T}\,U_{1i-1}$$
$$+ \frac{T}{C_2}\sum_{j=0}^{i-1}(U_{1j} - U_{2j}) + \frac{T}{C_1}\sum_{j=0}^{i-1}(U_{1j} - U_{gj}) = 0$$

$$(2\text{–loop}) \quad \left\{\frac{L_2 + L_3}{T} + (R_2 + R_3)\right\} U_{2i} - \frac{L_2 + L_3}{T}\,U_{2i-1}$$
$$+ \frac{T}{C_3}\sum_{j=0}^{i-1}(U_{2j} - U_{3j}) + \frac{T}{C_2}\sum_{j=0}^{i-1}(U_{2j} - U_{1j}) = 0$$

$$(3\text{–loop}) \quad \left\{\frac{L_3 + L_4}{T} + (R_3 + R_4)\right\} U_{3i} - \frac{L_3 + L_4}{T}\,U_{3i-1}$$
$$+ \frac{T}{C_4}\sum_{j=0}^{i-1}(U_{3j} - U_{Lj}) + \frac{T}{C_3}\sum_{j=0}^{i-1}(U_{3j} - U_{2j}) = 0$$

$$(4\text{–loop}) \quad \left\{\frac{L_4 + L_R}{T} + R_4\right\} U_{Li} - \frac{L_4 + L_R}{T}\,U_{Li-1}$$
$$- \frac{L_R}{T}\,(U_{Ri} - U_{Ri-1}) + \frac{T}{C_4}\sum_{j=0}^{i-1}(U_{Lj} - U_{3j}) = 0$$

$$(5\text{–loop}) \quad \frac{L_R}{T}\left\{(U_{Ri} - U_{Li}) - (U_{Ri-1} - U_{Li-1})\right\} + R_{Ri}U_{Ri} = 0,$$

where

$$R_{k1i} = \frac{0.19\rho}{A_{g1i-1}^2}, \; R_{k2i} = \left[\frac{0.5 - \dfrac{A_{g2i-1}}{A(1)}\left(1 - \dfrac{A_{g2i-1}}{A(1)}\right)}{A_{g2i-1}^2}\right]\rho,$$

$$L_{g1i} = \frac{\rho d_1}{A_{g1i-1}}, \; L_{g2i} = \frac{\rho d_2}{A_{g2i-1}}, \; R_{v1i} = 12\mu l_g^2\,\frac{d_1}{A_{g1i-1}^3},$$

$$R_{v2i} = 12 l_g^2\,\frac{d_2}{A_{g2i-1}^3}, \tag{17}$$

and

$$\frac{m_1}{T^2}(x_{1i} - 2x_{1i-1} + x_{1i-2}) + \frac{r_1}{T}(x_{1i} - x_{1i-1}) + s_1(x_{1i}) + k_c(x_{1i-1} - x_{2i-1}) = F_{1i}$$

$$\frac{m_2}{T^2}(x_{2i} - 2x_{2i-1} + x_{2i-2}) + \frac{r_2}{T}(x_{2i} - x_{2i-1}) + s_2(x_{2i}) + k_c(x_{2i-1} - x_{1i-1}) = F_{2i} ,$$

where

$$A_{g1i} = A_{g01} + 2l_g \cdot x_{1i} ,$$

$$A_{g2i} = A_{g02} + 2l_g \cdot x_{2i} ,$$

$$s_1(x_{1i}) = k_1 \cdot (x_{1i} + \eta_{k1} \cdot x_{1i-1}^3), \quad \text{for} \quad x_{1i} > -\frac{A_{g01}}{2l_g} ,$$

$$s_1(x_{1i}) = k_1 \cdot (x_{1i} + \eta_{k1} \cdot x_{1i-1}^3) + h_1 \cdot \left\{ \left(x_{1i} + \frac{A_{g01}}{2l_g} \right) + \eta_{h1} \cdot \left(x_{1i-1} + \frac{A_{g01}}{2l_g} \right)^3 \right\}, \quad \text{for} \quad x_{1i} \leqq -\frac{A_{g01}}{2l_g} ,$$

$$s_2(x_{2i}) = k_2 \cdot (x_{2i} + \eta_{k2} \cdot x_{2i-1}^3), \quad \text{for} \quad x_{2i} > -\frac{A_{g02}}{2l_g} ,$$

$$s_2(x_{2i}) = k_2 \cdot (x_{2i} + \eta_{k2} \cdot x_{2i-1}^3) + h_2 \cdot \left\{ \left(x_{2i} + \frac{A_{g02}}{2l_g} \right) + \eta_{h2} \cdot \left(x_{2i-1} + \frac{A_{g02}}{2l_g} \right)^3 \right\}, \quad \text{for} \quad x_{2i} \leqq -\frac{A_{g02}}{2l_g} ,$$

$$F_{1i}/d_1 l_g = P_{m1i} = P_s - 1.37 \frac{\rho}{2} \left(\frac{U_{gi}}{A_{g1i-1}} \right)^2 - \frac{1}{2} \left\{ R_{v1i} U_{gi} + \frac{L_{g1i}}{T} (U_{gi} - U_{gi-1}) \right\},$$

$$F_{2i}/d_2 l_g = P_{m2i} = P_{m1i} - \left\{ \frac{1}{2} (R_{v1i} + R_{v2i}) U_{gi} + (L_{g1i} + L_{g2i}) \frac{U_{gi} - U_{gi-1}}{T} \right\} - \frac{\rho}{2} U_{gi}^2 \left(\frac{1}{A_{g2i-1}^2} - \frac{1}{A_{g1i-1}^2} \right). \tag{18}$$

These difference equations were programmed in Fortran IV and compiled for experiment on one of the DDP-516 laboratory computers

of the Acoustics Research Department at Bell Laboratories.[13] Simultaneous solution of eqs. (17) and (18) yields all relevant volume velocities, glottal areas, and displacement. The time derivative of the mouth volume velocity (i.e., through the radiation load) is a good approximation to the radiated sound pressure.[12] Time samples of all dependent variables are obtained by iterating the solution for as many samples as are desired.

The sampling interval T is chosen as the longest interval that yields a stable solution to the difference equations. This interval is determined primarily by the time required for sound to transit the shortest length of the vocal tube. Because the distributed vocal tract is approximated as lumped constant T-sections, and because the behavior of these elements is further approximated by finite differences, the sampling interval T must be considerably shorter than the sound transit time through the shortest tube element. In the absence of appropriate sampling theory for this situation, the broad range of stable solutions was determined interactively on the DDP-516 computer and the longest stable interval used. In the present work, sampling rates in the range of 10 kHz to 30 kHz were used.

The iteration "loop" of the equations can be closed owing to the manner in which the glottal impedance elements and the forcing functions are taken to involve samples of glottal area; for example, current values of impedance and forcing function involve only past values of glottal areas. The iteration, therefore, proceeds as follows.

The cords and tract are initially assumed at rest, and initial currents are zero. The first sample of U_{gi} is calculated from loop-g using $A_{g,i-1} = A_{g0}$ (i.e., $x_{i-1} = 0$). The initial samples of all other loop currents are likewise calculated, out to the radiation load. The first sample of U_{gi} is then used to calculate the first samples of the forcing functions and, from the mechanical equations, the first samples of the displacements x_{1i} and x_{2i}. The latter dictate new values of A_{g1} and A_{g2} which are entered back into the glottal impedance elements for the calculation of the next sample of U_g and all other currents. The process is continued until as much of the solution as desired is obtained.

For synthesis of continuous speech, the vocal-tract area values change as do the values of P_s, A_{g0}, and cord constants.* These changes are slow by comparison to the sample variations in volume velocities, displacements, and pressure. The solutions for the continuously changing vocal system can therefore be considered as quasi-steady solutions of

* As indicated in Fig. 5, a cord-tension parameter, Q, constitutes an input to the vocal-cord model. This parameter determines the mechanical constants of the oscillator.

eqs. (17) and (18), and the mouth output samples taken as the synthetic speech signal.

IX. PHYSIOLOGICAL CONSTANTS FOR THE VOCAL-CORD MODEL

Few numerical values are available for the physiological parameters of the vocal cords. Using the sparse data available, the simulation on the DDP-516 computer was used to establish relevant ranges for the parameters.

First, the range of parameters which allows self-oscillation of the model was studied for a uniform vocal tract, 16 cm long, 5 cm^2 in cross-section, and terminated in the radiation load. The DDP-516 computer was used interactively to establish the self-oscillation region. The allowed oscillation range as a function of k_2 and k_c is shown in Fig. 6. In this plot, the axes are normalized by the factor m_1/m_2k_1. The parameters in the figure are the damping coefficients of the mechanical oscillators, ζ_1 and ζ_2. For these cases, all other glottal parameters

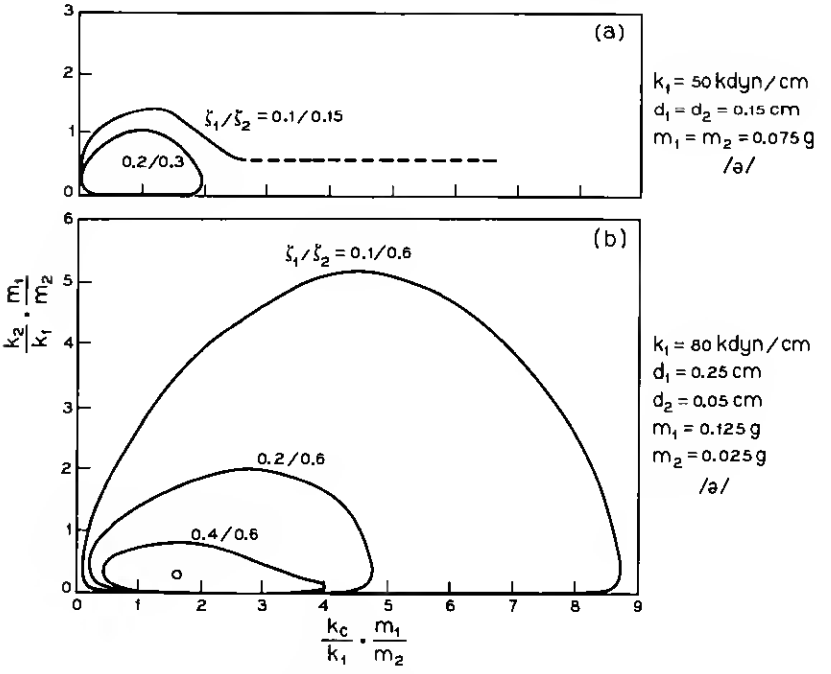

Fig. 6—Allowed regions of oscillation for the two-mass model. The parameter is the open-glottis damping ratio. The vocal-tract shape is for the vowel /ə/.

are held constant at physiologically realistic values: namely, $P_s = 8$ cm H_2O, $l_g = 1.4$ cm, $A_{g01} = A_{g02} = 0.05$ cm^2, thickness of the vocal cords $d_1 + d_2 = 0.3$ cm, total mass $m_1 + m_2 = 0.15$ g, nonlinear coefficient of the springs $\eta_{k1} = \eta_{k2} = 100$ and $\eta_{h1} = \eta_{h2} = 500$, and $h_1 = 3k_1$, $h_2 = 3k_2$. In particular, values for the spring constants are based upon measurements of static tensile stress versus displacement for an excised human larynx.[8] From these measurements, for example, η_k is deduced to be on the order of 50 to 100.

For Fig. 6a, the vocal cords are divided into equal parts, with thickness and mass 0.15 cm and 0.075 g, respectively, and with $k_t = 50$ kdyn/cm. For Fig. 6b, the lower part of the vocal-cord model is thicker than the upper part, that is, $d_1 = 0.25$ cm and $d_2 = 0.05$ cm, and the masses, $m_1 = 0.125$ g and $m_2 = 0.025$ g, are chosen proportional to the thicknesses, keeping the same total mass of 0.15 g as in Fig. 6a.

Kaneko[11] has measured the damped oscillations of a fresh excised human larynx when excited by a mechanical impulse and with no air flow through the glottis. From this data, the damping ratio for the excised human cords can be estimated to be of the order of 0.1 to 0.2 (which, incidentally, is the same order as deduced in the earlier simulations[1]). This range of damping seems particularly appropriate for the bulk of the cords, that is, for m_1 of the model.*

An acoustic load of the vocal tract, whose driving-point impedance has an inductive reactance at the fundamental frequency of the vocal-cord vibration, acts to enhance oscillation of the model. An increase in damping (loss) of the vocal tract at lower frequencies, as would be caused by wall vibration in the vocal tract, however, acts to oppose oscillation. Also, the tendency to oscillate is suppressed by an increase in the mechanical damping of m_1 and especially of m_2.

The behavior of the vocal-cord model, calculated for values of k_2 and k_c specified by the small circle in Fig. 6b, will now be considered. This glottal condition is chosen as the "typical" one throughout the present study; namely, $k_1 = 80$ kdyn/cm, $k_2 = 8$ dyn/cm, and $k_c = 25$ kdyn/cm.

* An equivalent damping ratio for the bulk of the cords can be estimated as follows:

$$(r_1 + r_2) = 2\,\zeta_1\sqrt{m_1k_1} + 2\,\zeta_2\sqrt{m_2k_2}.$$

For $k_c \to \infty$,

$$(r_1 + r_2) = 2\,\zeta_{equi}\sqrt{(m_1 + m_2)(k_1 + k_2)}.$$

Substituting (for the "typical" conditions) $m_2 = m_1/5$, $k_2 = k_1/10$, $\zeta_1 = 0.1$, and $\zeta_2 = 0.6$ gives

$$\zeta_{equi} = \frac{1}{\sqrt{66}}\,(\sqrt{50}\,\zeta_1 + \zeta_2) = 0.16,$$

which corresponds favorably with Kaneko's measurements.

X. RESULTS OF THE DIGITAL SIMULATION

The vocal-cord and vocal-tract program, specified by eqs. (17) and (18), was used interactively on the DDP-516 computer to calculate waveforms of glottal flow, glottal area, and mouth sound pressure.

10.1 *Waveforms for Typical Glottal Conditions*

Measurements made at the typical glottal condition and for a uniform vocal tract are illustrated in Fig. 7. Waveforms of A_{g1}, A_{g2}, U_g, and mouth sound pressure are shown for the initial 30 ms of voicing. The negative values of A_{g1} and A_{g2} indicate glottal closure. (One can imagine the cords forming into one another upon contact, and the negative areas correspond to the continued displacement of the center of mass of the cords.)

The results show that the phase difference between m_1 and m_2 is about 55 degrees, and the duty ratio (glottis open time to total period) is about 0.6. These values compare well with observations which have been made on human vocal cords by high-speed motion picture techniques[14] and by inverse filtering.[15] One notices the differences between the glottal area wave and the corresponding glottal flow wave, as pointed out in the earlier work.[1] The glottal flow wave is typically characterized by some temporal detail, asymmetry, and steep falling slope, while the area wave shows little temporal detail, is less steep,

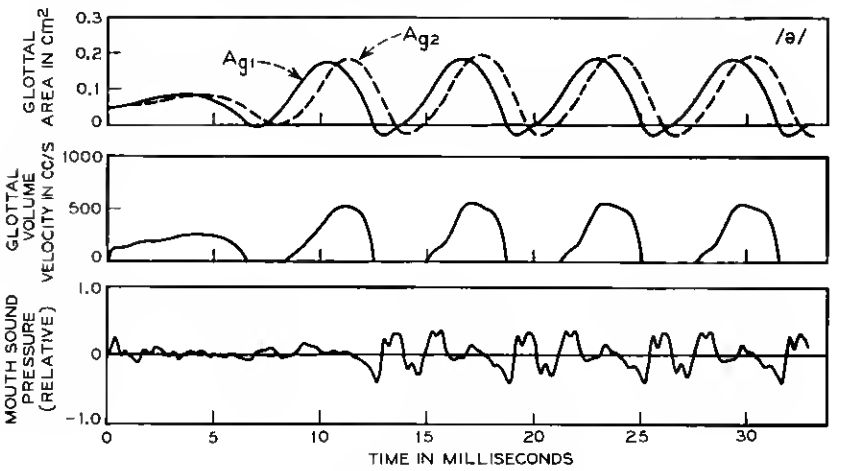

Fig. 7—Vocal-cord and vocal-tract functions computed from the DDP-516 simulation. Glottal areas, A_{g1} and A_{g2}, glottal volume velocity, U_g, and mouth-output sound pressure are shown for the initial 30 ms of voicing for the neutral vowel /ə/.

and tends to be more symmetrical. Because the cords are massive and are generally forced at a frequency above their natural frequency, their mechanical displacement does not reflect the detail of acoustic interaction which the glottal flow displays. The sound output wave reflects the periodicity established by the cord oscillator, and the greatest formant oscillation (or excitation) typically occurs (with about 0.5 ms delay) at the closing phase of the U_g wave. This effect has been observed previously.[2]

10.2 *Effect of Cord Stiffnesses*

The normalized k_2 versus k_c plane of Fig. 6 is a convenient medium for demonstrating the effects of spring constants. Using this plane, waveforms of U_g, A_{g1}, and A_{g2} are sketched for corresponding stiffnesses in Fig. 8. As before, tbe vocal tract is a uniform pipe (/ə/) and other glottal conditions are kept at their typical values.

An increase in k_c above the typical value reduces the phase difference between A_{g1} and A_{g2}. It also diminishes the steep falling slope of the flow waveform, and tends to make the wave more symmetrical and triangular. An increase in k_c also produces an increase in the build-up time required for the oscillation to settle to a steady state. For still larger values of k_c, close to the bounds of the oscillation range, both

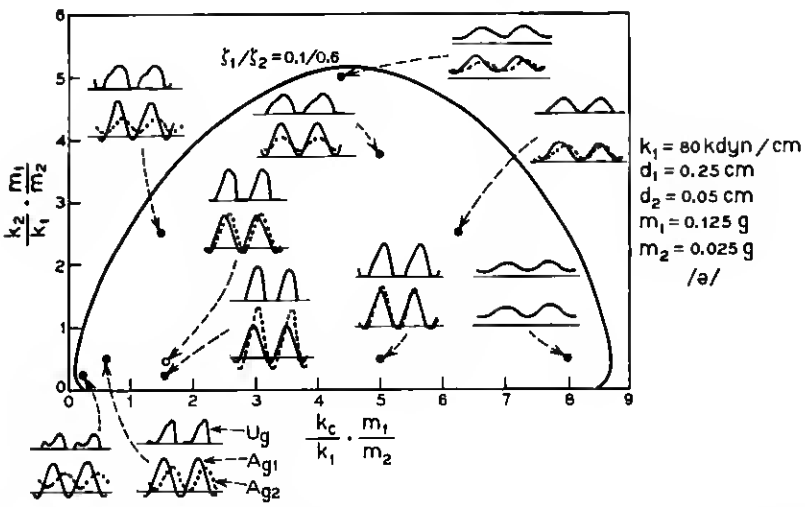

Fig. 8—Sketches of cord-tract functions for points on the k_2-k_c plane. The axes are normalized by the function (m_1/k_1m_2). The vowel is /ə/. Compare with Fig. 6b.

the glottal flow and area waveforms become sinusoidal on a dc component, and the glottis does not close.

The range of the sinusoidal behavior is expanded if the damping coefficients are made smaller. This special case is shown in Fig. 6a for the damping coefficients $\eta_{k1} = 0.1$ and $\eta_{k2} = 0.15$. Here, k_c has no limitation for the oscillation when k_2 is less than 20 kdyn/cm. Owing to the large k_c, the two-mass model behaves just as the one-mass model in the extended region in which tbe oscillation is sustained by the inductive reactance of the vocal tract and glottis. This projecting tail disappears with an increase in the losses, either of the vocal tract or of tbe vocal cords.

In contrast, an increase in k_2, with other conditions constant, decreases the amplitude of A_{g2} without a change of tbe phase difference. Further increase of k_2 leads to no closure of A_{g2} while A_{g1} can close completely during the cycle. Owing to the small amplitude of A_{g2} and its dc component, the glottal flow increases in upward roundness and also increases in duty ratio. A small, broad hump appears on the rising slope of the glottal flow wave, at which point the area A_{g1} is equal to A_{g2}.

By comparison, a decrease in k_2 increases the amplitude of A_{g2} and the glottal waves tend to a symmetrical form. This same dependence on k_2 and k_c also occurs for the case of equal thicknesses, $d_1 = d_2 = 0.15$ cm. A change in proportion of the damping coefficients, ζ_1 and ζ_2, also influences the relations between A_{g1} and A_{g2}. For example, the typical condition $\zeta_1 = 0.1$ and $\zeta_2 = 0.6$ produces an amplitude of A_{g1} slightly larger than that of A_{g2} for /ə/, as seen in Fig. 7. A smaller value of ζ_2 for the same values of ζ_1 and other parameters produces an amplitude of A_{g2} larger than A_{g1} without a change in phase difference. A steeper rising slope of the glottal area wave also results, but the falling slope remains unchanged.

10.3 *Effect of Neutral Area*

The behavior of the vocal-cord model with respect to the "phonation-neutral" area, or the equilibrium value A_{g0}, is another case where we can find correspondence between the complex behavior of the human vocal cords and the vibrations of the vocal-cord model. In human phonation the neutral area is maintained by laryngeal adjustment. Typical results from the simulation for different values of A_{g0} are illustrated in Fig. 9. These data were measured for the typical glottal conditions with $\zeta_1 = 0.1$ and $\zeta_2 = 0.6$ and for the vowel /i/. One sees that the build-up time required for the oscillation to reach a steady state increases as A_{g0} gets larger. The value $A_{g0} = 0.30$ cm^2 surpasses

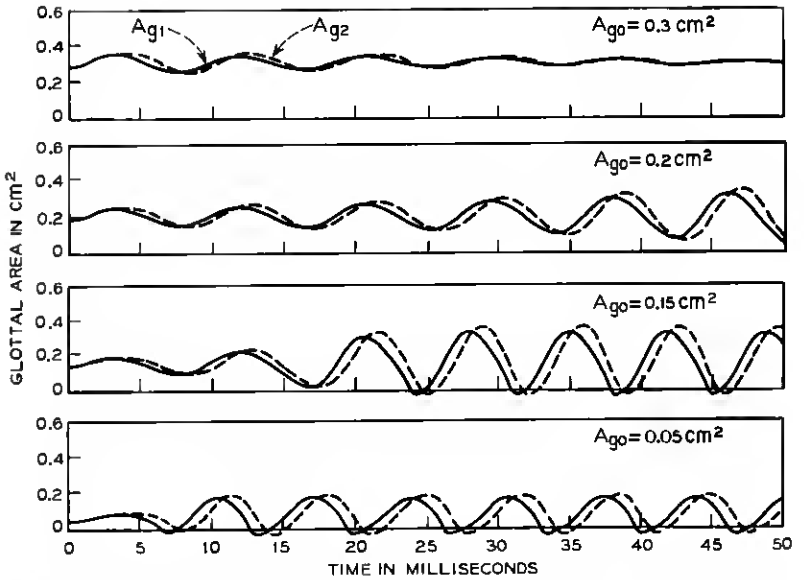

Fig. 9—Effect of the "phonation-neutral" or rest area, A_{g0}, upon the glottal area. The vowel is /i/.

a critical limit (about 0.25 cm^2) beyond which the model does not oscillate for these conditions.

During the voicing build-up time the pitch period is much longer than that of the steady-state oscillation. The change in pitch at the onset of voicing is similar to the starting motion of the human cords when they are brought to the phonation position from an open position. In this instance, unestablished low subglottal pressure also contributes to the reduction of the fundamental frequency. The oscillation period before cord closure is a value between the damped natural frequencies of the two mechanical oscillators. This is consistent with the value calculated from the acoustic theory of the two-mass model neglecting the collision and the nonlinearity of the springs.

Although the model, in this case, does not self-oscillate for $A_{g0} > 0.25$ cm^2, the maximum glottal area for phonation depends on the damping of the mechanical oscillators and of the vocal tract and on the subglottal pressure. For $\zeta_1 = 0.2$ and $\zeta_2 = 0.6$, and with $P_s = 8$ cm H_2O, the maximum glottal area reduces to about 0.20 cm^2. An increase in the phonation-neutral area also causes an increase in the amplitude of the vibration with no significant change in the period of the steady-state oscillation.

10.4 *Effect of Tract Shape*

Excitation of the vocal tract by the cord model was studied for several vowels. Area waves, glottal flow, and mouth-output sound pressure are shown for the vowels /i/, /u/, and /a/ in Figs. 10a, b, and c. For all these cases, the typical glottal conditions hold (same as for /ə/ in Fig. 7).

One notices that the waveforms of glottal area and the fundamental frequency are almost independent of the vocal-tract shape, while the shape can substantially influence the waveform of the glottal flow, similar to the results obtained from the one-mass model in the earlier work.[1] The acoustic interaction between the glottal flow and the acoustic load depends on the resonance characteristics of the vocal tract. Vowels having high resonant Q for the first formant show noticeable interaction in the glottal flow wave, as is seen for /a/. Also a low first formant can affect the glottal flow wave to a considerable extent, for example in /i/. However, the relatively large dissipation of the vocal tract in the frequency range of low first formants (such as for /i/ and /u/) caused primarily by vibration of the vocal-tract walls acts to reduce the interaction, but the glottal flow waveforms still differ markedly from each other. In all these cases, the tract losses are set to give bandwidths for the first formant equal to values measured on the human tract for the closed-glottis condition.[16]

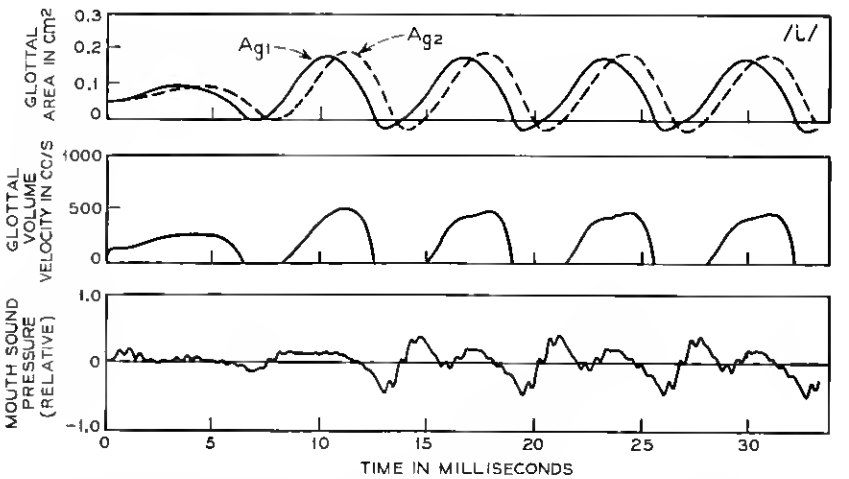

Fig. 10a—Results of the DDP-516 simulation for the vowel /i/ showing area waves, glottal flow, and mouth-output sound pressure.

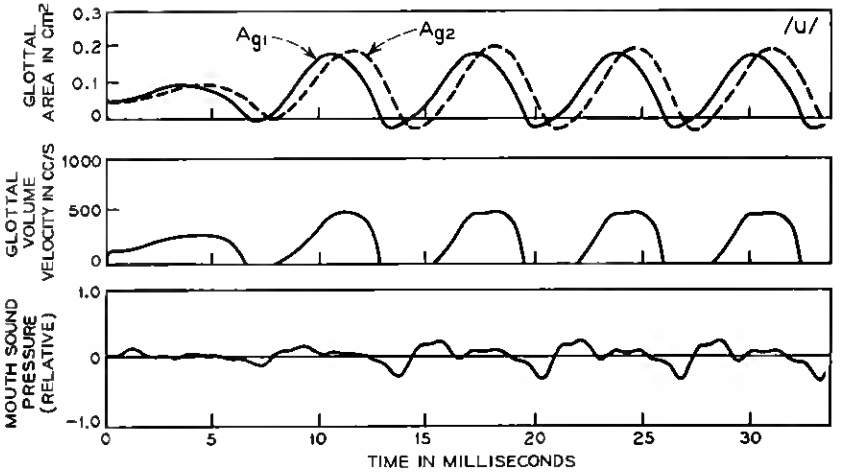

Fig. 10b—Same as Fig. 10a for the vowel /u/.

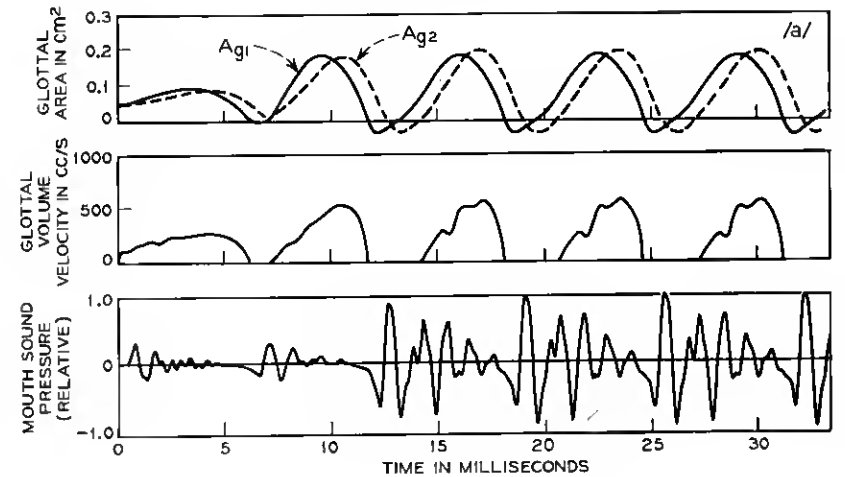

Fig. 10c—Same as Fig. 10a for the vowel /a/.

The data of Fig. 10* permit a comparison between the glottal waveform and the speech pressure wave. The comparison is familiar from the results of inverse filtering.[15,17] There is a delay time difference of about 0.5 ms between the waves, corresponding to the time required for the

* Sound spectrograms of the computed mouth-output sound pressure are shown for several vowels in Fig. 10d.

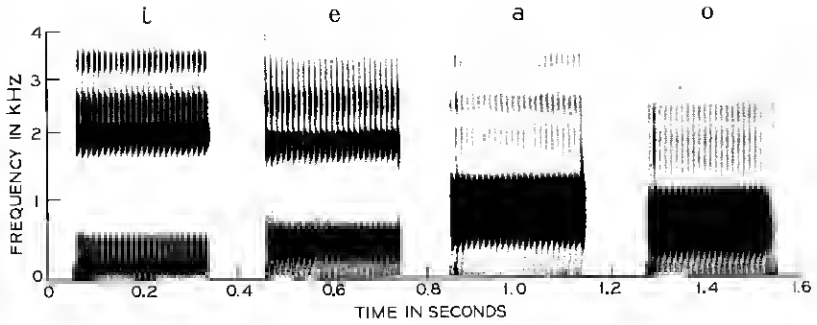

Fig. 10d—Sound spectrograms of the computed mouth-output sound pressure for the vowels /i, e, a, o/.

sound to travel from the glottis to the lips. The waveforms for /a/, /i/, and /u/ show that the formants are excited largely at the closure of the cords. The output pressure waves attenuate rapidly with increasing glottal area during the opening phase of the glottal cycle.

10.5 *Effect of Subglottal Pressure*

The influence of the subglottal pressure on the fundamental frequency of the vocal-cord vibration is another important aspect of voice production. Typical behavior of the model for these factors is shown in Fig. 11. The nonlinear coefficient of the spring, η_k, is shown as the parameter for the vowel /ə/. The data for the vowels /i/ and /a/ correspond to $\eta_k = 100$ solely. For all these cases, the coefficient describing the nonlinearity in the deformation of the vocal cords at closure is taken as $\eta_k = 5\eta_k$.

The slope of the fundamental frequency as a function of subglottal pressure is seen to be about 2.5 Hz/cm H_2O for $\eta_k = 100$, independent of the vowel configuration. This represents good agreement with measurements which have been made on human speech in the chest register by Hixon, et al.[18] The curve for $\eta_k = 0$, that is, linear springs, shows a saturation characteristic for subglottal pressures greater than 8 cm H_2O. These results suggest that pitch variations with subglottal pressure might be ascribed to two causes. One is the collision of the vocal cords at closure when the amplitude of vibration is not too large and the subglottal pressure is less than several cm H_2O. Another is the nonlinearity of the deflection of the muscles and ligaments at large amplitudes of vibration and at subglottal pressures more than several cm H_2O. In the latter case, the nonlinearity becomes dominant when large

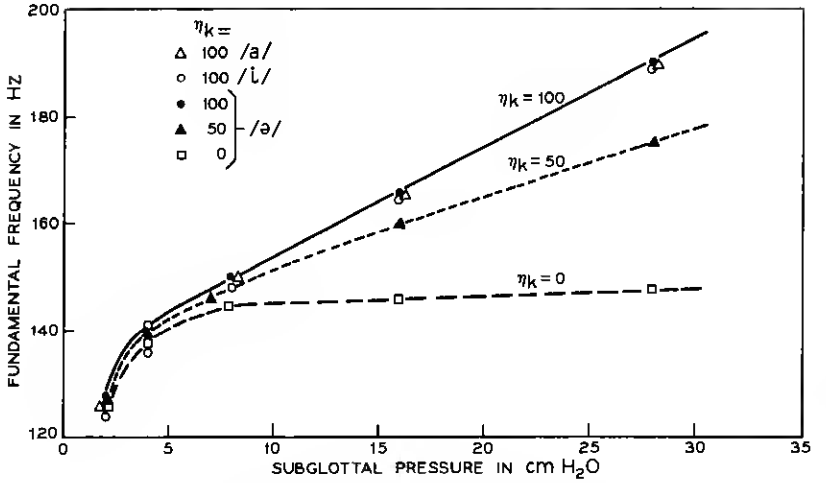

Fig. 11—Variation of fundamental frequency with subglottal pressure. The parameter is the nonlinear coefficient of the stiffnesses, η_k.

displacement amplitude increases the effective stiffness of the springs. This tends to increase fundamental frequency. The minimum subglottal pressure for vowel production is about 2 cm H_2O.

In the earlier work with the one-mass model, significant influences were found on fundamental frequency as a function of tract configuration. This influence was due in large part to the pressure recovery assumed at the glottal outlet, namely $1/2\,P_B$ according to van den Berg's data. When the intraglottal pressure distribution derived here is used in the one-mass model, the interaction across vowels and with subglottal pressure is much less.

The two-mass model becomes a one-mass model if k_c is increased to a large value. For this condition, the variation in fundamental frequency with subglottal pressure is shown for several vowels in Fig. 12. The behavior is similar to the two-mass model. Under these conditions, the duty ratio of the former tends to be slightly greater than the latter.

Duty ratio is another aspect of the model that can be compared to human phonation. An increase in subglottal pressure produces an increase in glottal flow and in glottal amplitude. Duty ratio (open time to total period) decreases for this increase in subglottal pressure and is asymptotic to a value around 0.5, as shown in Fig. 13. This value compares well with measurements on natural speech.[12]

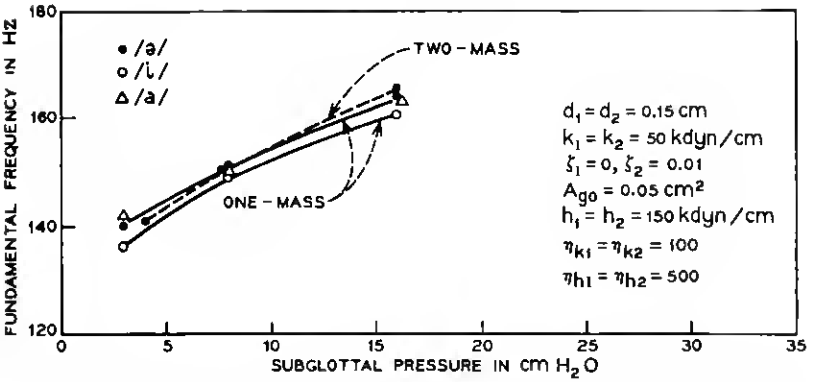

Fig. 12—Variations of fundamental frequency with subglottal pressure for the one-mass and two-mass models. The parameter is vowel configuration.

10.6 *Effect of Cord Tension*

As in the previous studies,[1-3] it is convenient to apply a "tension parameter," Q, to control fundamental frequency. This can be achieved by causing the masses and thicknesses to be scaled down and the springs scaled up by the factor Q, causing the fundamental frequency to vary proportionally with Q. Phase difference, duty ratio, and glottal area waveforms are essentially uninfluenced by Q, and the amplitudes of glottal area and glottal flow decrease gradually with increasing Q. The glottal flow waveform also varies in detail depending on the fundamental frequency, because the formants contributing to the temporal detail of the glottal flow are unchanged while the period of the glottal

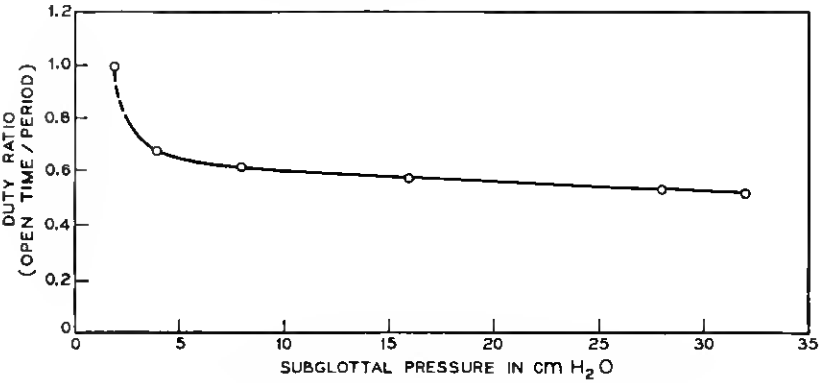

Fig. 13—Variation of duty ratio with subglottal pressure.

flow varies as a function of Q. Changes in flow waveform with pitch variation are greatest in cases where acoustic interaction is especially pronounced (such as for /a/).

In human speech the duty ratio has a tendency to increase with the fundamental frequency.[19] This feature can also be given to the vocal-cord model by modifying the coupling-tension parameter k_c to increase more than in linear proportion to Q. A variation as Q^2 appears more realistic. Physiologically this corresponds to the considerable decrease in compliance and thickness of the vocal cords when they are stretched by contraction of the cricothyroid muscle and other muscles associated with contracting of the vocalis. The increase of k_c more than proportional to Q is equivalent to shifting the glottal operation condition on a line parallel to the abscissa in Fig. 6. As indicated in Fig. 8, a shift to the right reduces the phase difference and increases the duty ratio without changing other features of the cord vibration, except near the boundaries of the oscillation range.

Behavior of the cord model with the Q parameter so defined is shown in Figs. 14 and 15. Variations in waveforms with Q are shown for the vowel /a/ in Fig. 14. The relations between fundamental frequency, duty ratio, and amplitude of glottal area with Q are plotted in Fig. 15. Variation of the duty ratio with frequency falls into the range measured in inverse filtering experiments.[19]

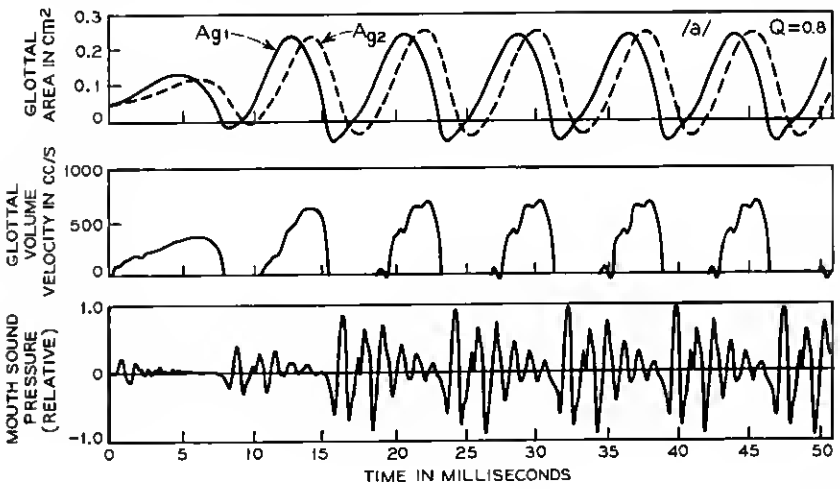

Fig. 14a—Effect of tension parameter, Q, on cord-tract output for the vowel /a/. $Q = 0.8$.

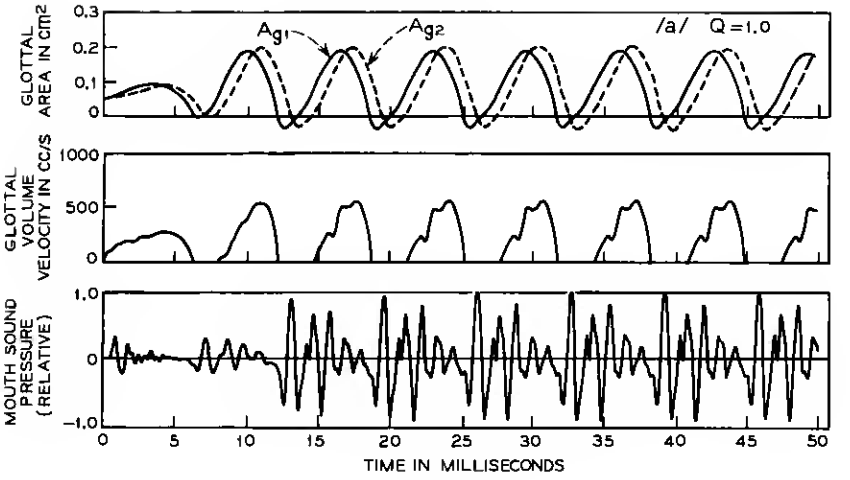

Fig. 14b—Same as Fig. 14a with $Q = 1.0$.

XI. INTERACTION EFFECTS WITH LARGE ACOUSTIC LOADS

11.1 *Differences Between Two-Mass and One-Mass Models*

The measurements discussed previously show that the fundamental frequency and the area waveforms of the cord model are not strongly influenced by tract geometry. The interaction with glottal flow, however,

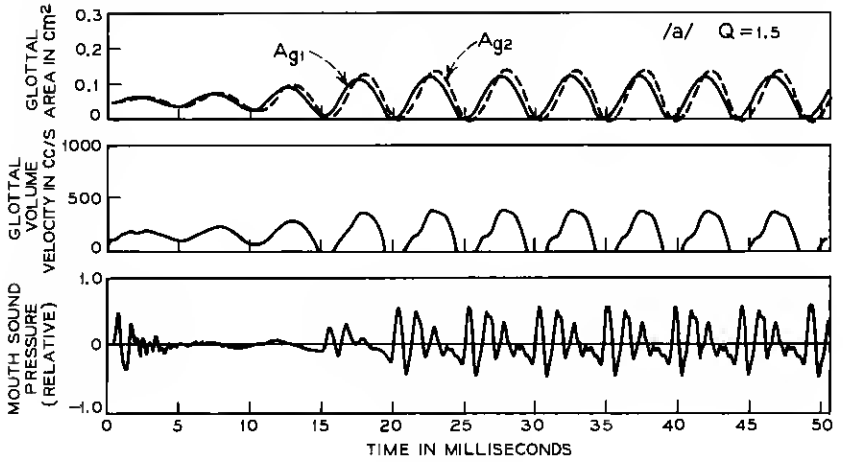

Fig. 14c—Same as Fig. 14a with $Q = 1.5$.

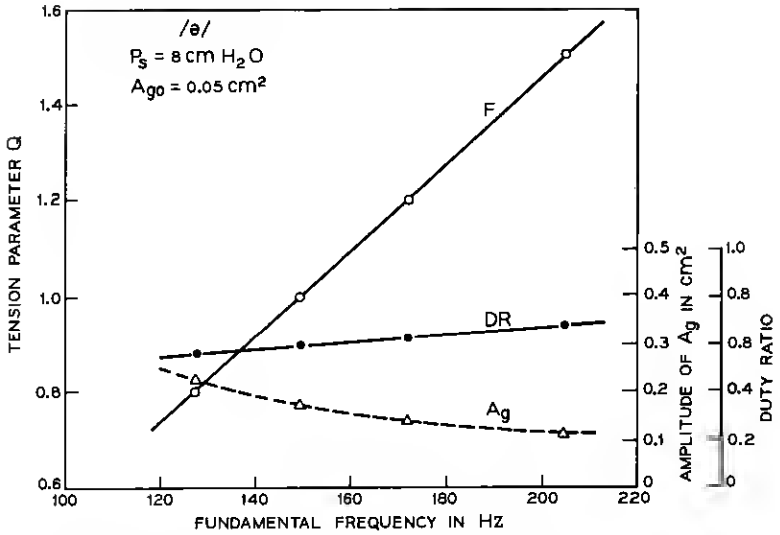

Fig. 15—Effect of tension parameter (Q) on fundamental frequency (F), duty ratio (DR), and glottal area (A_g).

is marked. We have further investigated the effect of acoustic load by lowering the frequency of the lowest resonance of the acoustic load (the first formant) into the range of the fundamental frequency. This increases the driving-point impedance at the fundamental frequency and strong coupling between source and load is expected.

The formant frequencies are lowered by lengthening the simulated vocal tract. Measurements of the fundamental frequency are shown in Fig. 16 as a function of the length of a uniform vocal-tract tube, 5 cm^2 in cross-section. Data are shown for both the two-mass cord model and an equivalent one-mass model ($k_c \rightarrow \infty$). The measurements are for the typical glottal conditions. The shunt impedance of the vocal-tract wall (wall vibration) is not taken into account *per se*, and this effect is only approximated by an increase in damping for the first 16-cm section of the tube (as was used for the /ə/ configuration). The remaining tube is regarded as an ideal hard-wall tube. The first resonance frequency of the vocal-tract tube, F_{01}, is shown by the solid line.

The frequency of the two-mass model decreases more gradually than that of the one-mass model with increasing the tube length. When the oscillation frequency of the former meets the first formant frequency of the vocal-tract tube, a sharp increase of the fundamental frequency occurs for further increase in tube length. The frequency returns to

almost the same value as for a short tube. The frequency jump occurs at the resonant frequency of the vocal-tract tube, independent of dissipation and of glottal conditions. For example, an increase in acoustic dissipation of the vocal tract and a decrease in mechanical damping of m_1 and m_2 raises the onset frequency of the jump, but the frequency where the jump occurs is still the first resonant frequency of the tube. The variation of frequency with vocal-tube length is shown for two conditions of damping in Fig. 16.

The curve of F_{01} as a function of tube length marks the boundary between an inductive driving-point impedance (to the left) and a capacitive driving-point impedance (to the right). The frequency jump for the two-mass model, which occurs at F_{01} regardless of the glottal conditions, places its new oscillation in the capacitive region, that is, between the first pole and second zero of the driving-point impedance.

A frequency jump also occurs in the one-mass model. In this case, however, the jump is to the original frequency for which the driving-point impedance is still an inductive impedance, that is, between the

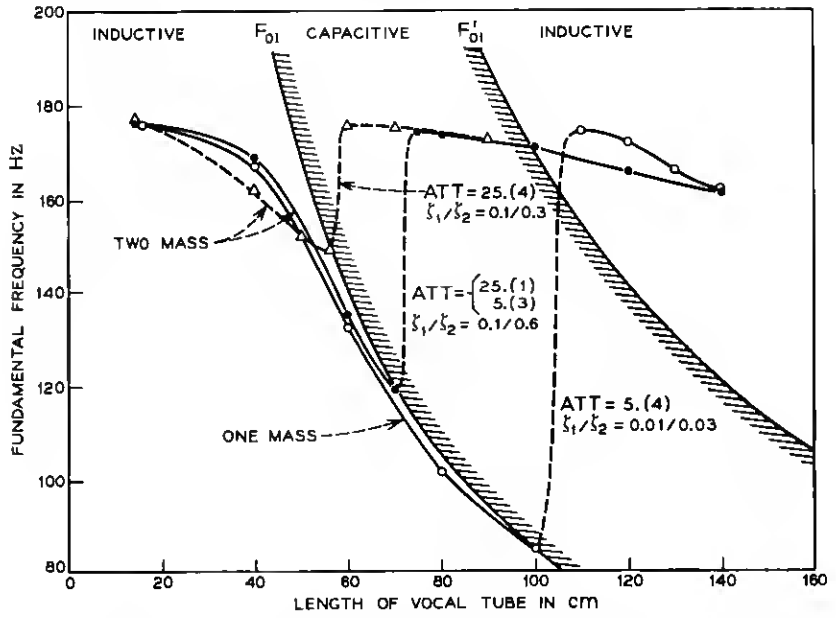

Fig. 16—Variation of fundamental frequency with acoustic load for the two-mass and one-mass models. F_{01} shows the frequency of the first pole of the driving-point impedance, and F_{01}' shows the first zero.

second zero and second pole of the driving-point impedance. This behavior can be predicted by an analysis of the oscillator with a uniform transmission line as a load.

11.2 *Effects of Acoustic Load on Human Voicing*

For comparison with the model behavior, we have measured similar loading effects in human voicing. To bring a first formant resonance into the range of the voice pitch, subjects phonated into a long metal tube the length of which was periodically changed from 39 cm to 73 cm by a motor (i.e., a bazooka-like sliding pipe). The subjects were instructed to pronounce the sustained vowel /ə/ at medium sound level and with constant glottal adjustment regardless of the change in tube length. Fundamental frequency (pitch) measurements were made at several frequencies in the chest register. Typical results for one subject are shown in Fig. 17.

The voice pitch was measured at 10-ms intervals by a pitch-extracting program.[20] The length of the metal tube (exclusive of the subject's vocal tract) is also indicated on the abscissa along with the corresponding time scale for the length change. Adjacent open and closed points (circles or triangles) pertain to different cycles of the pipe in one set of measurements. One sees frequency jumps similar to those in the two-mass model. However, the observed onset frequencies of the jumps are generally higher than the resonant frequency of the compound tube consisting of the metal tube and the subject's vocal tract (neglecting the shunt impedance of the vocal-tract wall). The deviation from the resonant frequency becomes especially noticeable for lower frequencies.

Toward an interpretation, it is known that the shunting impedance caused by vibration of the walls of the vocal tract produces a "cutoff frequency" of the sound transmission and constrains the lowest first formant frequency of the vocal tract.[21] This effect will contribute to raising the resonant frequency of the compound tube in a frequency range near the cutoff frequency. In the present instance, one could conceive of the walls of the cheeks, pharynx, and soft velum to yield to vibration because of the vocal-tract geometry for /ə/ and because of the long wavelength. At the cutoff frequency of the vocal tract, the first resonance frequency of the combined vocal tract and metal pipe is essentially that of the metal pipe alone. The latter is shown in Fig. 17 by the broken line.

From Fig. 17, we can presume the cutoff frequency of the vocal tract for /ə/ to be a little lower than 200 Hz. The effect of the wall vibration could thus account for the rightward shift of the observed pitch jumps. The rightward shift is most noticeable at the lower frequencies as this

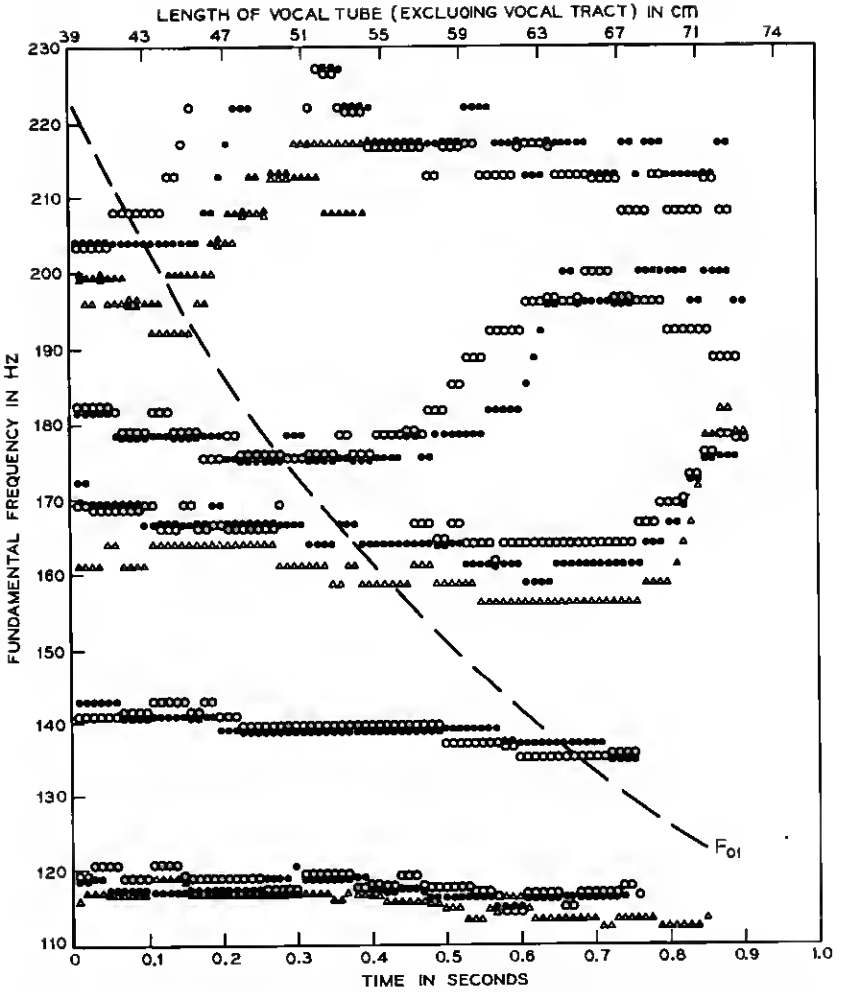

Fig. 17—Fundamental frequency measurements made for a human subject when the acoustic load on his vocal cords is varied. The acoustic load is varied by periodically changing the length of a uniform tube fitted to the subjects' mouth. The broken line shows the first resonant (pole) frequency of the uniform tube.

argument would predict. Even with these uncertainties, we see the close similarity in the dependence of fundamental frequency on acoustic load between the human larynx and the two-mass model.* It is further of interest that the vocal cords can self-oscillate without regenerative

* *Note added in proof:* After this paper was written, we measured the "cutoff frequency" for the vocal tract and tube combination. We found its value to be 195 Hz.

feedback from the subglottal and supraglottal system. In addition, the vibration of the soft walls of the vocal tract acts as a buffer to aid stable operation in the presence of coupling between the vocal cords and the vocal tract as the latter takes on a wide variety of shapes.

XII. CONCLUSION

The two-mass formulation of the vocal-cord model is seen to yield physiologically realistic behavior. In particular, the phase differences between upper and lower cord-edges corresponds well with motion observed in high-speed photography. The two-mass formulation also leads to a natural correlate to chest and falsetto register with coupling stiffness (lax in chest and tense in falsetto) being an important factor along with mass and thickness of the cords.

The computer measurements show that the two-mass model is capable of oscillation just above the resonant frequencies of the acoustic load (i.e., the formant frequencies of the vocal tract), duplicating a capability of the human cords. The one-mass model cannot oscillate in this frequency range, where the driving-point reactance is capacitive. Further, the intra-glottal pressure distribution derived for use with the two-mass model yields cord-tract interaction similar to human speech. Fundamental frequency varies with subglottal pressure approximately as 2 to 3 Hz/cm H_2O, and changes in vowel configuration do not markedly influence the fundamental frequency. Closures tighter than those which occur in vowel shapes (for example, at consonant-vowel boundaries) can of course influence the fundamental frequency. The improved intra-glottal pressure distribution is also applicable to a one-mass formulation, and it produces physiologically realistic cord-tract interactions with a one-mass model.

The programmed cord oscillator and the digitally simulated vocal tract constitute a complete synthesizer for voiced sounds. The system so implemented has potential for speech synthesis applications such as computer voice response. Especially for techniques such as text synthesis,[22] the cord model and vocal tract offer means for natural control of tract and larynx parameters, i.e., subglottal pressure, cord tension, neutral area, and tract shape. These parameters appear sufficient for describing both voiced and voiceless sounds in continuous speech.[3] In some synthesis applications, the complexity of the two-mass model may not be needed and a simpler one-mass formulation may serve. In normal voice production, phonation occurs at a fundamental frequency always below the first vocal resonance (formant). Here, the driving-point imped-

ance is inductive and the one-mass oscillator performs acceptably, particularly with the improved intra-glottal pressure distribution.

The two-mass model, because of its physiological detail, also provides a potential tool for medical analyses of voice disorder. Although the present simulation assumes bilateral symmetry of the opposing cords, asymmetric configurations can be implemented. The effects of deficiencies such as unilateral cord paralysis can therefore be investigated and quantified. Biomedical engineering is making increased use of digital simulations of physiological behavior. The simulation technique described here not only permits acoustic analysis of voice functions but of human respiration as well.

XIII. ACKNOWLEDGMENTS

We wish to thank several members of the Acoustics Research Department for their substantial contributions to this study. A. E. Rosenberg collaborated on the design and measurements with the impedance tube, D. E. Dudgeon made early simulations with the two-mass program, D. E. Bock assisted in an interactive implementation of the program on the DDP-516 computer, and K. Shipley programmed the pitch extractor used for the real voice measurements.

REFERENCES

1. Flanagan, J. L., and Landgraf, L. L., "Self-Oscillating Source for Vocal-Tract Synthesizers," Proc. IEEE-AFCRL Symposium on Speech Commun. and Process., Boston, Mass., (Nov. 1967), also published in IEEE Trans. Audio and Electroacoustics, *AU-16*, (March 1968), pp. 57–64.
2. Flanagan, J. L., "Use of an Interactive Laboratory Computer to Study an Acoustic-Oscillator Model of the Vocal Cords," IEEE Trans. Audio and Electroacoustics, *AU-17*, (March 1969), pp. 2–6.
3. Flanagan, J. L., and Cherry, L., "Excitation of Vocal-Tract Synthesizers," J. Acoust. Soc. Amer., *45*, (March 1969), pp. 764–769.
4. Dudgeon, D. E., "Two-Mass Model of the Vocal Cords," J. Acoust. Soc. Amer., *48*, (July 1970), p. 118A.
5. Ishizaka, K., "On Models of the Larynx," J. Acoust. Soc. Japan, *22*, 1966, pp. 293–294.
6. Ishizaka, K., and Matsudaira, M., "What Makes the Vocal Cords Vibrate," 6th Int. Congr. Acoust., Tokyo, (Aug. 1968), pp. B1–3.
7. Ishizaka, K., and Matsudaira, M., "Acoustic Theory of a Two-Mass Model of the Vocal Cords," unpublished work.
8. Ishizaka, K., and Kaneko, T., "On Equivalent Mechanical Constants of the Vocal Cords," J. Acoust. Soc. Japan, *24*, No. 5, 1968, pp. 312–313.
9. Kaufmann, W., *Fluid Mechanics*, New York: McGraw-Hill Co., 1963, p. 111.
10. van den Berg, J. W., Zantema, J. T., and Doornenbal, Jr., P., "On the Air Resistance and the Bernoulli Effect of the Human Larynx," J. Acoust. Soc. Amer., *29*, 1957, pp. 626–631.
11. Ishizaka, K., Kaneko, T., and Matsudaira, M., unpublished work.
12. Flanagan, J. L., *Speech Analysis, Synthesis and Perception*, 2nd Edition, New York, Berlin: Springer Verlag, 1972.

13. Flanagan, J. L., "Focal Points in Speech Communication Research," Proc. 7th Int. Cong. Acoust., Budapest, (August 1971). Also, IEEE Trans. Commun. Tech., (December 1971).
14. Farnsworth, D. W., "High-Speed Motion Pictures of the Human Vocal Cords," Bell Labs Record, *18*, No. 7 (March 1940), pp. 203–208.
15. Miller, R. L., "Nature of the Vocal Cord Wave," J. Acoust. Soc. Amer., *31*, 1969, pp. 667–677.
16. Fujimura, O., and Lindqvist, J., "Sweep Tone Measurements of Vocal-Tract Characteristics," J. Acoust. Soc. Amer., *49*, (Feb. 1971), pp. 554–558.
17. Holmes, J. N., "An Investigation of the Volume Velocity Waveform at the Larynx During Speech by Means of an Inverse Filter," Proc. Speech Commun. Seminar, Stockholm, 1962.
18. Hixon, T. J., Mead, J., and Klatt, D. H., "Influence of Forced Transglottal Pressure Changes on Vocal Fundamental Frequency," J. Acoust. Soc. Amer., *49*, (Jan. 1971), p. 105A.
19. Lindqvist, J., "The Voice Source Studied by Means of Inverse Filtering," Quarterly Report, Speech Transmicro Laboratory, Stockholm, Sweden, (Jan. 1970).
20. Gold, B., and Rabiner L., "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain," J. Acoust. Soc. Amer., *46*, (Aug. 1969), pp. 442–448.
21. Fant, G., "Speech at High Ambient Air-Pressure," STL-QPSR, (Feb. 1964), pp. 9–21.
22. Flanagan, J. L., Coker, C. H., Rabiner, L. R., Schafer, R. W., and Uneda, N., "Synthetic Voices for Computers," IEEE SPECTRUM, *7*, (October 1970), pp. 22–45.